Assignment #3: Database implementation

1. Specifications

You will implement ER diagrams from previous phase using either **MySQL, PostgreSQL or SQLite** as a backend database engine. Your implementation must be in **Python3 or Java programming language**. You will need to implement the structure of the database (tables with relations), provide sample data for your tables and provide implementation of SELECT queries (see below for details). Please note that the implementation may require you to iterate over your ER design and fine-tune it.

2. Goals

2.1 Create database structure: convert your ER diagram into a real database using Data Definition Language (DDL). Keep in mind that your database structure must also preserve integrity and consistency of your data and must be in the third normal form (normalization will be presented in future classes, tutorials and labs).

2.2 Populate your database with sample data: provide sample data for each table (HINT: writing a Python3 script or Java to generate data may save you some time).

2.3 Implement SELECT queries: write a Python3 script or Java that will interact with your database and perform select queries (see below for list of queries).

2.4 Design an interface to access the queries via the programming language of your choice.

3. SELECT Queries

Below you will find ten tasks, each describes a scenario of interaction with your database. You will need to implement them using SQL's SELECT statement. You can replace four of them with any other queries of your interest and motivate your choice.

HINT: Your ER diagram might require some modifications in order to allow these tasks. When you replace four of the tasks it is advisable to replace those requiring most modifications with some requiring little or no modifications.

HINT: We do not require you to implement all tasks using pure SQL solutions (i.e. you may combine SQL and programming language processing). However, more points will be awarded to SQL solutions.

3.1 A customer claims she forgot her bag in a car and asks to help. She was using cars several times this day, but she believes the right car was red and its plate starts with "AN". Find all possible cars that match the description.

3.2 Company management wants to get a statistics on the efficiency of charging stations utilization. Given a date, compute how many sockets were occupied each hour.

Example INPUT:
01/11/2018

00h-01h: 0		12h-13h: 4	
01h-02h: 1		13h-14h: 1	
02h-03h: 1		14h-15h: 3	
03h-04h: 2		15h-16h: 0	
04h-05h: 0		16h-17h: 3	
05h-06h: 0		17h-18h: 2	
06h-07h: 3		18h-19h: 0	
07h-08h: 3		19h-20h: 0	
08h-09h: 4		20h-21h: 4	
09h-10h: 4		21h-22h: 1	
10h-11h: 6		22h-23h: 0	
11h-12h: 7		23h-24h: 0	

3.3 Company management considers using price increasing coefficients. They need to gather statistics for one week on how many cars are busy (% to the total amount of taxis) during the morning (7AM - 10 AM), afternoon (12AM - 2PM) and evening (5PM - 7PM) time.

Example OUTPUT:

Morning	Afternoon	Evening
80	37	72

3.4 A customer claims that he was charged twice for the trip, but he can't say exactly what day it happened (he deleted notification from his phone and he is too lazy to ask the bank), so you need to check all his payments for the last month to be be sure that nothing was doubled.

3.5 The department of development has requested the following statistics:
- Average distance a car has to travel per day to customer's order location
- Average trip duration
Given a date as an input, compute the statistics above.

3.6 In order to accommodate traveling demand, the company decided to distribute cars according to demand locations. Your task is to compute top-3 most popular pick-up locations and travel destination for each time of day: morning (7am-10am), afternoon (12am-2pm) and evening (5pm-7pm).

3.7 Despite the wise management, the company is going through hard times and can't afford anymore to maintain the current amount of self-driving cars. The management decided to stop using 10% of all self-driving cars, which take least amount of orders for the last 3 months.

3.8 The company management decided to participate in the research on "does customer location of residence depend on how many charging station the self-driving cars was using the same

day". Now you as DB developer need to provide this data. You've decided to collect the data for each day within one month and then sum them up.

EXAMPLE:

A customer with ID 234 has used the taxi three times during the last month. These were 2 different taxis. First of them visited the charging station twice the same day of the first trip and once at the date of the second customer's trip. The second taxi used charging stations 2 times the same day when the customer ordered it, so the total is 5.

Example INPUT:
starting date (e.g 1.10.2017)

Example OUTPUT:

UserId	Amount
234	5
423	2
623	13

3.9 The company management decided to optimize repair costs by buying parts in bulks from providers for every workshop. Help them decide which parts are used the most every week by every workshop and compute the necessary amount of parts to order.

EXAMPLE:

First workshop most often requires windshield washer motors (about 10 every week on average). Second workshop most often requires sparking cables (about 7 every week on average).

3.10
The company management decided to cut costs by getting rid of the most expensive car to maintain. Find out which car type has had the highest average (per day) cost of repairs and charging (combined).

4. Deliverables
4.1 An SQL dump file with sample data and database structure.
4.2 A Python3 script or Java code of implementation of SELECT queries.

5. Grading
70% of the grade will be assigned to your schema and Python3 or Java interface implementation.
30% of the grade will be assigned to the implementation of selected queries

6. Submissions
Submission is in Moodle, only **one person per team** should submit.
Make sure that your submission includes names of all team members.